

Напишете интерактивна програма, която за дадени откъси от текст, взети от Wikipedia (виж примера по-долу), определя езика на който е написан всеки от текстовете. След всяко определяне, вашата програма получава верния отговор, така че програмата да може да се самообучава и да дава по-добри отговори при следващи извиквания.

Езиците са номерирани с числата от 0 до 55. Всеки откъс от текст съдържа точно 100 знака, зададени като масив E от 100 цели числа, всяко от 1 до 65 535, включително.

Напишете процедура **excerpt(E)**, където E е масива от 100 числа, представящ текста. Процедурата **excerpt(E)** трябва да извика **language(L)** само веднъж, където L е номера на езика, според вашата програма. Функцията **language(L)** е реализирана от грейдера и връща като резултат правилния номер на езика.

Грейдерът вика **excerpt(E)** 10 000 пъти. Вашият резултат е равен на отношението на броя на текстовете, за които вашата програма е познала езика към общия брой текстове.

Може да използвате различни методи за решаване на задачата. Методът на *Rocchio* е подход, който дава верен резултат в около 40% от случаите. Методът пресмята подобие между текущия текст E и всеки език L от езиците, разглеждани досега и избира най-близкия език. Подобие се дефинира като общия брой на различните знаци от текста E , които се срещат в някой от предишните текстове от езика L .

Пример. Вж. английския текст!

Файлът на грейдера `grader.in.1` съдържа 10 000 такива текстове. Тези 56 езика са майчините езици на участниците в IOI 2010. Текстовете са взети от началото на първите параграфи от случайно избрани статии на 56 езика от Wikipedia.

Subtask 1 [30 points]

Вашият резултат трябва да е 0.3 или по-добър от този на грейдера.

Subtask 2 [up to 80 points]

Вашият резултат трябва да са $114(\alpha - 0.3)$, закръглено до най-близкото цяло, където α е вашият резултат.

Имплементация: вж. английския текст!